Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 6682–6686, October 1987
Biochemistry

Brain cDNA clone for human cholinesterase

(butyrylcholinesterase/pseudocholinesterase)

CHARLES MCTIERNAN*, STEVE ADKINS†, ARNAUD CHATONNET†, THERESA A. VAUGHAN†, CYNTHIA F. BARTELS†, MATTHEW KOTT†, TERRONE L. ROSENBERRY*, BERT N. LA DU†, AND OKSANA LOCKRIDGE†‡

*Pharmacology Department, Case Western Reserve University School of Medicine, Cleveland, OH 44106; and †Pharmacology Department, Medical Science Building 1, University of Michigan Medical School, Ann Arbor, MI 48109-0626

Communicated by Minor J. Coon, June 8, 1987

ABSTRACT **A cDNA library from human basal ganglia was screened with oligonucleotide probes corresponding to portions of the amino acid sequence of human serum cholinesterase (EC 3.1.1.8). Five overlapping clones, representing 2.4 kilobases, were isolated. The sequenced cDNA contained 207 base pairs of coding sequence 5′ to the amino terminus of the mature protein in which there were four ATG translation start sites in the same reading frame as the protein. Only the ATG coding for Met-(−28) lay within a favorable consensus sequence for functional initiators. There were 1722 base pairs of coding sequence corresponding to the protein found circulating in human serum. The amino acid sequence deduced from the cDNA exactly matched the 574 amino acid sequence of human serum cholinesterase, as previously determined by Edman degradation. Therefore, our clones represented cholinesterase (EC 3.1.1.8) rather than acetylcholinesterase (EC 3.1.1.7). It was concluded that the amino acid sequences of cholinesterase from two different tissues, human brain and human serum, were identical. Hybridization of genomic DNA blots suggested that a single gene, or very few genes, coded for cholinesterase.**

cDNA clones for acetylcholinesterase from *Torpedo californica* (1), *Drosophila melanogaster* (2) and *Torpedo marmorata* (3) have been isolated and sequenced. No other cholinesterases have been cloned, though a preliminary report regarding cholinesterase cDNA from human brain has appeared (4). One of our laboratories has recently completed the amino acid sequence of human serum cholinesterase (acylcholine acylhydrolase, EC 3.1.1.8) (5). These sequencing results allowed us to prepare oligonucleotide probes, which we have used to identify cDNA clones for cholinesterase.

Cholinesterase and acetylcholinesterase (EC 3.1.1.7) are generally considered to be products of different genes (6). The evidence includes the observation that there is no immunological crossreaction between them (7, 8). Furthermore, there are known genetic variants of human cholinesterase, including the silent variant with zero activity in serum (9), but no corresponding genetic variants of acetylcholinesterase. The function of cholinesterase is unknown, while the function of acetylcholinesterase is to terminate nerve impulse transmission at cholinergic synapses. Cholinesterase and acetylcholinesterase are highly similar, but not identical, with regard to substrate preference, the types of chemicals that inhibit activity, tissue distribution, and occurrence as globular and asymmetric molecular forms (6). Comparison of amino acid sequences, glycosylation patterns, and disulfide bonding is available to date only for human cholinesterase and acetylcholinesterase from fish or fly, as the two enzyme types have not been sequenced or cloned from a single species. Despite the species difference, human cholinesterase and *Torpedo* acetylcholinesterase are 54% identical in amino acid sequence (5). They have nearly the same number, 574 and 575, of amino acids per catalytic subunit, and their differences in molecular weight are due to differences in glycosylation, with acetylcholinesterase having four and cholinesterase having nine carbohydrate chains per subunit (1, 5, 10, 11). Their disulfide bonds are located in precisely the same locations, suggesting similar protein folding (10, 11). The present report shows the nucleotide sequence of human cholinesterase cDNA,[§] a result that should be useful for future comparisons aimed at determining the structural relationship between cholinesterase and acetylcholinesterase.

METHODS

cDNA Library Screening. A human cDNA library from the basal ganglia of a 1-day-old brain was kindly provided by R. A. Lazzarini (12). This library is now available through the American Type Culture Collection (Rockville, MD). The cDNA was cloned into bacteriophage λgt11 via *Eco*RI linkers. The library contains $\approx 10^6$ independent recombinants. Screening was carried out by the *in situ* plaque-hybridization method of Benton and Davis (13) as modified by Maniatis *et al.* (14). Table 1 lists the oligonucleotide probes used for finding the first three positive clones. Oligonucleotides 5 and 6 were synthesized by the University of Michigan DNA-synthesizing facility. All other oligonucleotides were synthesized by P-L Biochemicals. The oligonucleotides were labeled with [γ-^{32}P]ATP, using phage T4 polynucleotide kinase. Prehybridization as well as hybridization was in 0.25% Carnation instant non-fat dry milk (15) dissolved in 6× SSC (6× SSC is 0.9 M NaCl/0.09 M sodium citrate, pH 7.0). Hybridization was for 12–20 hr at 36°C for all oligonucleotides except no. 5, which was hybridized at 46°C. Posthybridization washes were at room temperature in 6× SSC/0.05% pyrophosphate, followed by a 10-min wash at 41–43°C in 6× SSC/0.05% pyrophosphate for probes 3, 4, 7, 8, 9, and 10. More stringent conditions were used for the final wash for probes 5 and 6. Probe 5 was washed for 10 min at 46°C in 2× SSC/0.1% $NaDodSO_4$. Probe 6 was washed for 10 min at 46°C in 6× SSC/0.05% pyrophosphate.

A 130-base-pair (bp) fragment of clone Z3 was used as a probe to find the overlapping clone Z2. The 130-bp fragment had been produced by digestion with *Sau*3A1. It was subcloned into phage M13, where it was radiolabeled by synthesizing a labeled complement strand. The double-stranded fragment was released from M13 by digestion with *Bam*HI

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‡To whom reprint requests should be addressed.

§This sequence is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) under accession no. J02964.

Table 1. Oligodeoxynucleotide probes

Probe	Protein sequence		Probe sequence† (5′ to 3′)
	Amino acids*	Location	
1	WKNQFN	557–562	TGGAARAAYCARTTYAA
2	MMDWKN	554–559	TTYTTCCARTCCATCAT
3	WNNYMMDWKN	550–559	TTITTCCAITCCATCATITAITTITTCCA
4	AEWEWKAGFH	539–548	TGIAAICCIGCITTCCAITCCCAITCIGC
5	WPEWMGVMHGYEIEFVFG	430–447	CCAAAGACAAATTCAATTTCATATCCATGCATGACTCCCATCCATTCAGGCCA
6	EWGNNAFFYYFEH	411–423	ATGTTCAAAATAATAAAAAAAAGCATTATTTCCCCATTC
7	KEFQEG	348–353	AARGARTTYCARGARGG
8	VYGAPGF	331–337	AAICCIGGIGCICCITAIAC
9	GVNKDEGTAF	320–329	AAIGCIGTICCITCITCITTITTIACICC
10	GQFKKTQI	310–317	ATITGIGTITTITTIAAITGICC

*Single-letter code for amino acids: A, alanine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; M, methionine; N, asparagine; P, proline; Q, glutamine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

†R = A or G; Y = C or T; I, deoxyinosine.

and *Eco*RI. It was purified by polyacrylamide gel electrophoresis and electroeluted into a dialysis bag. Approximately 10^5 plaques were screened by hybridization at 60°C, followed by three brief washes at room temperature in 2× SSC/0.1% $NaDodSO_4$ and a 10-min wash at 59°C in 0.1× SSC/0.2% $NaDodSO_4$.

A 270-bp fragment of clone Z2 was used as a probe to find overlapping clone Z13. The 270-bp fragment was produced by digesting clone Z2 with *Bam*HI and *Eco*RI. The *Eco*RI site was in the linker joining human DNA to the λ arm and was not in the cholinesterase cDNA. The 270-bp fragment was amplified in plasmid pUC18. The 270-bp fragment was isolated on low-melting-point agarose and radiolabeled by the random oligolabeling method of Feinberg and Vogelstein (16). Hybridization was at 68°C. Posthybridization washes were with 0.1× SSC/0.5% $NaDodSO_4$. Three brief washes at room temperature were followed by a 120-min wash at 68°C.

DNA Sequencing. Sequencing was by the dideoxy method of Sanger *et al.* (17). Restriction fragments were cloned into M13mp18 and M13mp19. Primers for the sequencing reaction included the universal primer, oligonucleotides 1 and 2 in Table 1, and λgt11 primers.

Computer Analysis. The Protein Identification Resource contains computer programs that allowed us to store and align our DNA sequence files and to compare the DNA sequences of our clones to those in the computer data bank.¶

RESULTS

The oligonucleotides listed in Table 1 were synthesized for use as probes for human cholinesterase cDNA. One of our laboratories had previously determined the complete amino acid sequence of human serum cholinesterase by Edman degradation of overlapping peptides (5). The probes correspond to regions of the amino acid sequence representing minimal codon ambiguity. Our most successful oligonucleotide probe proved to be the unique-sequence 53-mer (probe 5) because it could be used at the highest stringency. The first clone, OH57, was found by using probe 5 alone. The 53-mer probe matched the cDNA clone in 50 out of 53 nucleotides.

Clone Z3 (Fig. 1) was found by screening 1.5×10^5 plaques of the human brain cDNA library after making four nitrocellulose "lifts" from each plate. One filter was hybridized with the 53-mer (probe 5); one with the 39-mer (probe 6); one with a mixture of probes 4, 8, and 10; and one filter was hybridized with a mixture of probes 3, 7, and 9. Clone Z3 was positive on all four filters. Clone Z35 was found by hybridizing duplicate filters with two different probes, nos. 3 and 4. Overlapping clones that completed the coding sequence at the 5′ end were found by using fragments of clones Z3 and Z2 as probes. The 130-bp fragment of Z3 indicated in Fig. 1 was used to screen 10^5 plaques, which led to the isolation of clone Z2. The 270-bp fragment of Z2 was used to screen 2×10^6 plaques, which led to isolation of clone Z13.

The location of one intron was suggested from the sequencing results of clone Z35. This clone had 105 extra nucleotides at its 5′ end in addition to the 614 nucleotides indicated in Fig. 1. The extra 105 nucleotides did not match the known amino acid sequence and were absent in clone Z3 as determined by DNA sequencing. The 420 bp of Z35 immediately downstream from these 105 nucleotides matched clone Z3 in all but 2 nucleotides. The junction between the 105 nucleotides and the rest of clone Z35 contained the sequence TTCCATATTTTACAGG, which is similar to the consensus sequence present at the 3′ splice junction of an intron (18). Thus, it appears that the 105 nucleotides belong to an intron, and that this intron is located between nucleotides 1600 and 1601 in Fig. 2. We considered the possibility that these 105 extra nucleotides represented alternative splicing leading to a

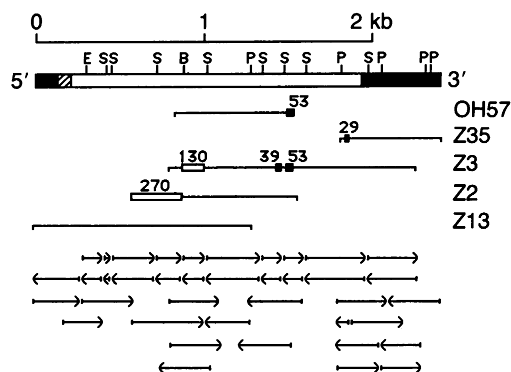

FIG. 1. Restriction map and sequencing strategy for cholinesterase cDNA. Scale at top shows length in kilobases (kb). The open bar represents the 1722 nucleotides that code for the 574 amino acids of mature cholinesterase. The hatched bar represents 84 nucleotides of the signal-peptide region. The solid bar at the 3′ end represents 492 nucleotides of the 3′ untranslated sequence. Restriction sites used for sequencing are indicated by E (*Eco*RI), S (*Sau*3A1), B (*Bam*HI), and P (*Ssp* I). Lines immediately below the restriction map show the sizes and designations for five cDNA inserts in λgt11. The solid boxes are oligonucleotide probes 53, 29, and 39 nucleotides in length. They are shown on the clones they identified. The open boxes are cDNA probes, 130 and 270 bp in length, prepared from clones Z3 and Z2, respectively. Lines with arrowheads show the direction and extent of DNA sequencing.

¶Protein Identification Resource (1987) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC), Release 11.

```
    *** Met Ser Val Gln Ser Asn Leu Gln Ala Gly Ala Ala Ala Ala Ser Cys Ile Ser  -52
 AC TGA ATG TCA GTG CAG TCC AAT TTA CAG GCT GGA GCA GCA GCT GCA TCC TGC ATT TCC -154

Pro Lys Tyr Tyr Met Ile Phe Thr Pro Cys Lys Leu Tyr His Leu Cys Cys Arg Glu Ser  -32
CCG AAG TAT TAC ATG ATT TTC ACT CCT TGC AAA CTT TAC CAT CTT TGT TGC AGA GAA TCG  -94
            -28
Glu Ile Asn Met His Ser Lys Val Thr Ile Ile Cys Ile Arg Phe Leu Phe Trp Phe Leu  -12
GAA ATC AAT ATG CAT AGC AAA GTC ACA ATC ATA TGC ATC AGA TTT CTC TTT TGG TTT CTT  -34
                                          -1  +1
Leu Leu Cys Met Leu Ile Gly Lys Ser His Thr.Glu Asp Asp Ile Ile Ile Ala Thr Lys    9
TTG CTC TGC ATG CTT ATT GGG AAG TCA CAT ACT.GAA GAT GAC ATC ATA ATT GCA ACA AAG   27

Asn Gly Lys Val Arg Gly Met Asn Leu Thr Val Phe Gly Gly Thr Val Thr Ala Phe Leu   29
AAT GGA AAA GTC AGA GGG ATG AAC TTG ACA GTT TTT GGT GGC ACG GTA ACA GCC TTT CTT   87

Gly Ile Pro Tyr Ala Gln Pro Pro Leu Gly Arg Leu Arg Phe Lys Lys Pro Gln Ser Leu   49
GGA ATT CCC TAT GCA CAG CCA CCT CTT GGT AGA CTT CGA TTC AAA AAG CCA CAG TCT CTG  147

Thr Lys Trp Ser Asp Ile Trp Asn Ala Thr Lys Tyr Ala Asn Ser Cys Cys Gln Asn Ile   69
ACC AAG TGG TCT GAT ATT TGG AAT GCC ACA AAA TAT GCA AAT TCT TGC TGT CAG AAC ATA  207

Asp Gln Ser Phe Pro Gly Phe His Gly Ser Glu Met Trp Asn Pro Asn Thr Asp Leu Ser   89
GAT CAA AGT TTT CCA GGC TTC CAT GGA TCA GAG ATG TGG AAC CCA AAC ACT GAC CTC AGT  267

Glu Asp Cys Leu Tyr Leu Asn Val Trp Ile Pro Ala Pro Lys Pro Lys Asn Ala Thr Val  109
GAA GAC TGT TTA TAT CTA AAT GTA TGG ATT CCA GCA CCT AAA CCA AAA AAT GCC ACT GTA  327

Leu Ile Trp Ile Tyr Gly Gly Gly Phe Gln Thr Gly Thr Ser Ser Leu His Val Tyr Asp  129
TTG ATA TGG ATT TAT GGT GGT GGT TTT CAA ACT GGA ACA TCA TCT TTA CAT GTT TAT GAT  387

Gly Lys Phe Leu Ala Arg Val Glu Arg Val Ile Val Val Ser Met Asn Tyr Arg Val Gly  149
GGC AAG TTT CTG GCT CGG GTT GAA AGA GTT ATT GTA GTG TCA ATG AAC TAT AGG GTG GGT  447

Ala Leu Gly Phe Leu Ala Leu Pro Gly Asn Pro Glu Ala Pro Gly Asn Met Gly Leu Phe  169
GCC CTA GGA TTC TTA GCT TTG CCA GGA AAT CCT GAG GCT CCA GGG AAC ATG GGT TTA TTT  507

Asp Gln Gln Leu Ala Leu Gln Trp Val Gln Lys Asn Ile Ala Ala Phe Gly Gly Asn Pro  189
GAT CAA CAG TTG GCT CTT CAG TGG GTT CAA AAA AAT ATA GCA GCC TTT GGT GGA AAT CCT  567
                                 ...
Lys Ser Val Thr Leu Phe Gly Glu SER Ala Gly Ala Ala Ser Val Ser Leu His Leu Leu  209
AAA AGT GTA ACT CTC TTT GGA GAA AGT GCA GGA GCA GCT TCA GTT AGC CTG CAT TTG CTT  627

Ser Pro Gly Ser His Ser Leu Phe Thr Arg Ala Ile Leu Gln Ser Gly Ser Phe Asn Ala  229
TCT CCT GGA AGC CAT TCA TTG TTC ACC AGA GCC ATT CTG CAA AGT GGA TCC TTT AAT GCT  687

Pro Trp Ala Val Thr Ser Leu Tyr Glu Ala Arg Asn Arg Thr Leu Asn Leu Ala Lys Leu  249
CCT TGG GCG GTA ACA TCT CTT TAT GAA GCT AGG AAC AGA ACG TTG AAC TTA GCT AAA TTG  747

Thr Gly Cys Ser Arg Glu Asn Glu Thr Glu Ile Ile Lys Cys Leu Arg Asn Lys Asp Pro  269
ACT GGT TGC TCT AGA GAG AAT GAG ACT GAA ATA ATC AAG TGT CTT AGA AAT AAA GAT CCC  807

Gln Glu Ile Leu Leu Asn Glu Ala Phe Val Val Pro Tyr Gly Thr Pro Leu Ser Val Asn  289
CAA GAA ATT CTT CTG AAT GAA GCA TTT GTT GTC CCC TAT GGG ACT CCT TTG TCA GTA AAC  867

Phe Gly Pro Thr Val Asp Gly Asp Phe Leu Thr Asp Met Pro Asp Ile Leu Leu Glu Leu  309
TTT GGT CCG ACC GTG GAT GGT GAT TTT CTC ACT GAC ATG CCA GAC ATA TTA CTT GAA CTT  927

Gly Gln Phe Lys Lys Thr Gln Ile Leu Val Gly Val Asn Lys Asp Glu Gly Thr Ala Phe  329
GGA CAA TTT AAA AAA ACC CAG ATT TTG GTG GGT GTT AAT AAA GAT GAA GGG ACA GCT TTT  987

Leu Val Tyr Gly Ala Pro Gly Phe Ser Lys Asp Asn Asn Ser Ile Ile Thr Arg Lys Glu  349
TTA GTC TAT GGT GCT CCT GGC TTC AGC AAA GAT AAC AAT AGT ATC ATA ACT AGA AAA GAA 1047

Phe Gln Glu Gly Leu Lys Ile Phe Phe Pro Gly Val Ser Glu Phe Gly Lys Glu Ser Ile  369
TTT CAG GAA GGT TTA AAA ATA TTT TTT CCA GGA GTG AGT GAG TTT GGA AAG GAA TCC ATC 1107

Leu Phe His Tyr Thr Asp Trp Val Asp Asp Gln Arg Pro Glu Asn Tyr Arg Glu Ala Leu  389
CTT TTT CAT TAC ACA GAC TGG GTA GAT GAT CAG AGA CCT GAA AAC TAC CGT GAG GCC TTG 1167

Gly Asp Val Val Gly Asp Tyr Asn Phe Ile Cys Pro Ala Leu Glu Phe Thr Lys Lys Phe  409
GGT GAT GTT GTT GGG GAT TAT AAT TTC ATA TGC CCT GCC TTG GAG TTC ACC AAG AAG TTC 1227

Ser Glu Trp Gly Asn Asn Ala Phe Phe Tyr Tyr Phe Glu His Arg Ser Ser Lys Leu Pro  429
TCA GAA TGG GGA AAT AAT GCC TTT TTC TAC TAT TTT GAA CAC CGA TCC TCC AAA CTT CCG 1287

Trp Pro Glu Trp Met Gly Val Met His Gly Tyr Glu Ile Glu Phe Val Phe Gly Leu Pro  449
TGG CCA GAA TGG ATG GGA GTG ATG CAT GGC TAT GAA ATT GAA TTT GTC TTT GGT TTA CCT 1347

Leu Glu Arg Arg Asp Asn Tyr Thr Lys Ala Glu Glu Ile Leu Ser Arg Ser Ile Val Lys  469
CTG GAA AGA AGA GAT AAT TAC ACA AAA GCC GAG GAA ATT TTG AGT AGA TCC ATA GTG AAA 1407

Arg Trp Ala Asn Phe Ala Lys Tyr Gly Asn Pro Asn Glu Thr Gln Asn Asn Ser Thr Ser  489
CGT TGG GCA AAT TTT GCA AAA TAT GGG AAT CCA AAT GAG ACT CAG AAC AAT AGC ACA AGC 1467

Trp Pro Val Phe Lys Ser Thr Glu Gln Lys Tyr Leu Thr Leu Asn Thr Glu Ser Thr Arg  509
TGG CCT GTC TTC AAA AGC ACT GAA CAA AAA TAT CTA ACC TTG AAT ACA GAG TCA ACA AGA 1527

Ile Met Thr Lys Leu Arg Ala Gln Gln Cys Arg Phe Trp Thr Ser Phe Phe Pro Lys Val  529
ATA ATG ACG AAA CTA CGT GCT CAA CAA TGT CGA TTC TGG ACA TCA TTT TTT CCA AAA GTC 1587

Leu Glu Met Thr Gly Asn Ile Asp Glu Ala Glu Trp Glu Trp Lys Ala Gly Phe His Arg  549
TTG GAA ATG ACA GGA AAT ATT GAT GAA GCA GAA TGG GAG TGG AAA GCA GGA TTC CAT CGC 1647

Trp Asn Asn Tyr Met Met Asp Trp Lys Asn Gln Phe Asn Asp Tyr Thr Ser Lys Lys Glu  569
TGG AAC AAT TAC ATG ATG GAC TGG AAA AAT CAA TTT AAC GAT TAC ACT AGC AAG AAA GAA 1707

Ser Cys Val Gly Leu ***                                                          574
AGT TGT GTG GGT CTC TAA TTAATAGATTTACCCTTTATAGAACATATTTTCCTTTAGATCAAGGCAAAAATAT 1780

CAGGAGCTTTTTTACACACCTACTAAAAAAGTTATTATGTAGCTGAAACAAAAATGCCAGAAGGATAATATTGATTCCT 1859

CACATCTTTAACTTAGTATTTTACCTAGCATTTCAAAACCCAAATGGCTAGAACATGTTTAATTAAATTTCACAATATA 1938
                                                                 ......
AAGTTCTACAGTTAATTATGTGCATATTAAAACAATGGCCTGGTTCAATTTCTTTCTTTCCTTaataaaTTTAAGTTTT 2017

TTCCCCCCAAAATTATCAGTGCTCTGCTTTTAGTCACGTGTATTTTCATTACCACTCGTAAAAAGGTATCTTTTTTAAA 2096

TGAATTAAATATTGAAACACTGTACACCATAGTTTACAATATTATGTTTCCTAATTAAAATAAGAATTGAATGTCAATA 2175

TGAGATATTAAAATAAGCACAGAAAATCAAAAAAAAAAA                                         2214
```

FIG. 2. *(Legend appears at the bottom of the opposite page.)*

cholinesterase with an extra or different peptide near its carboxyl terminus. However, there were five stop codons within the 105 nucleotides, a result compatible with an intron but not with a coding region. It is not uncommon to find unspliced introns or possible cloning artifacts in cDNA clones (3, 19). We have not ruled out the interpretation that these 105 nucleotides may be a cloning artifact rather than an intron.

The nucleotide sequence is shown in Fig. 2. The cholinesterase found in human serum corresponds to nucleotides 1–1722 and contains 574 amino acids. The amino acid sequence derived from the nucleotide sequence of the brain cDNA is a perfect match with the amino acid sequence of cholinesterase found in human serum. This confirms the amino acid sequence determined by Edman degradation (5), including those regions where an overlap had not been obtained but depended on homology with *Torpedo* acetylcholinesterase. Of special interest was the finding that asparagine occurred at residues 17, 57, 106, 241, 256, 341, 455, 481, and 486. It had been reported (5) that these nine asparagines were glycosylated. Since asparagine cannot be detected by amino acid composition analysis due to hydrolysis of the amide by HCl, and since glycosylated asparagine is not detected as a phenylthiohydantoin amino acid derivative, the identification of asparagine had been based on indirect evidence. The nucleotide sequence shown here supports the conclusion that these residues are asparagines and that they are linked to carbohydrate chains.

Blot hybridization of human placental DNA that had been digested with various restriction enzymes yielded the results shown in Fig. 3. The enzymes and the DNA fragment sizes were *Eco*RI (2.45 kb), *Hin*dIII (4.95 kb), *Bam*HI (no fragment), *Kpn* I (no fragment), *Pst* I (14 kb), *Eco*RI/*Bam*HI (1.9 kb), *Bam*HI/*Hin*dIII (3.1 and 1.85 kb), *Eco*RI/*Hin*dIII (2.45 kb), *Eco*RI/*Xba* I (1.9 kb), and *Eco*RI/*Pst* I (2.45 kb). These results indicate that there may be a single copy or very few copies of the cholinesterase gene.

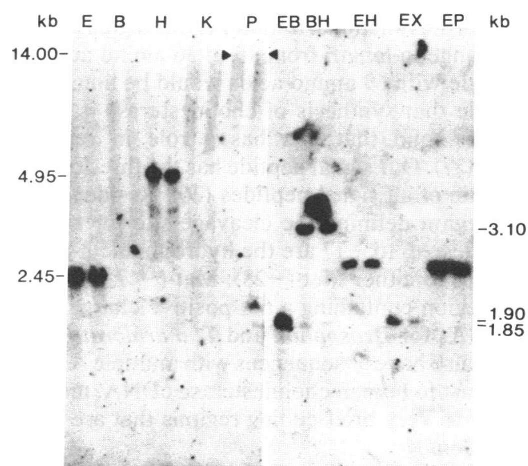

FIG. 3. Hybridization of genomic blots with ^{32}P-oligolabeled cDNA clone OH57. Each lane contained 25 μg of human placental genomic DNA digested with *Eco*RI (E), *Bam*HI (B), *Hin*dIII (H), *Kpn* I (K), *Pst* I (P), or combinations of these (EX, *Eco*RI plus *Xba* I). Electrophoresis in 1.2% agarose was followed by DNA transfer to nitrocellulose filter. Hybridization was at 68°C for 24 hr with probe (4 ng/ml) in 6× SSC/0.25% dry milk. The blots were washed in 2× SSC/0.1% $NaDodSO_4$ at 68°C (3 times, 1 hr each) and exposed for 20 hr with intensifying screen.

DISCUSSION

The coding region of the cDNA clone isolated from human brain is a perfect match with the amino acid sequence of human serum cholinesterase (5). Therefore we conclude that the amino acid sequences of cholinesterase from two different human tissues, brain and serum, are identical. The cholinesterase present in human serum is thought to be synthesized in the liver (20). The cholinesterase in serum is the globular tetrameric G4 form. Human fetal and adult brain contain G4 and G1 forms of cholinesterase, with G4 being the most abundant (21, 22). It is not known whether G4 and G1 are encoded by the same or different mRNAs, and it is not known which form is represented by our cDNA from human brain.

To date the evidence leads us to expect that cholinesterase and acetylcholinesterase will be shown to have different primary amino acid sequences. The amino terminus of acetylcholinesterase from human erythrocytes, Glu-Gly-Ala-Glu-Asp-Ala- (23), is different from the Glu-Asp-Asp-Ile-Ile-Ile- for human serum cholinesterase. The carboxyl termini for these two enzymes are also different, being -His-Gly for erythrocyte acetylcholinesterase (24) and -Gly-Leu for serum cholinesterase. The 18 amino acid sequence represented by our 53-mer probe is highly conserved in human cholinesterase and in *Torpedo and Drosophila* acetylcholinesterases. This suggests that these same 18 amino acids are likely to be conserved in human acetylcholinesterase. Acetylcholinesterase is more abundant than cholinesterase in human brain, regardless of developmental stage (21, 22). For these reasons, we expected to find acetylcholinesterase as well as cholinesterase cDNA by using our 53-mer oligonucleotide as a probe. However, we have found no acetylcholinesterase candidate clones that differ from the sequence in Fig. 2.

Two possible genetic variations were found in clone Z35. Base 1615 was adenine in clone Z35 but was guanine in clone Z3. Base 1914 was guanine in Z35 but was adenine in Z3. Fig. 2 shows guanine at nucleotide 1615, coding for Ala-539, because amino acid sequencing had shown Ala-539. An adenine at nucleotide 1615 would code for threonine. A rare variant would not have been detected by amino acid sequencing because the cholinesterase protein had been purified from pooled human serum. The cDNA library was from mRNA of a single person. If the donor was heterozygous, the cDNA library would contain an equal number of clones representing the rare and normal variants. The possibility that these nucleotide differences are cloning artifacts must also be considered.

Upstream of the beginning of the mature protein, in the signal-peptide region, are 207 nucleotides coding for 69 amino acids. There are four ATG triplets in the same reading frame as the protein, and one ATG triplet in a different reading frame. In 95% of all cases the ATG triplet nearest the 5′ end of the mRNA is the initiator signal (25). This would suggest that the ATG at position −207, corresponding to Met-(−69), could be the codon that initiates translation. However, this ATG does not lie in a consensus sequence for initiator signals. Only the ATG coding for Met-(−28) lies within a favorable consensus sequence, AATATGC, commonly found in functional initiators (25). Another reason favoring

FIG. 2 (*on opposite page*). Nucleotide sequence and amino acid sequence of human cholinesterase. Nucleotide 1 is the G of the GAA codon for the amino-terminal amino acid (Glu, amino acid 1). The active-site serine is SER-198. The polyadenylylation signal (aataaa) is shown in lowercase letters in the 3′ untranslated region. The nucleotide locations of clones used to construct this sequence are as follows: Z13, −213 to 1066; Z2, 400 to 1381; Z3, 583 to 2027; OH57, 625 to 1341; and Z35, 1600 to 2214. Nucleotide 1615 (Ala-539) was G in clone Z3 but was A in clone Z35. Nucleotide 1914 was A in clone Z3 but was G in clone Z35.

Met-(−28) as the functional initiator is that signal peptides are known to range in length from 15 to 36 amino acids (26). A signal peptide with 69 amino acids would be unusually long. It is possible that synthesis of cholinesterase is initiated at several sites, and that this has a role in regulation of expression (27). Our signal peptide has the three components characteristic of all signal peptides (26). Residues −1 to −5 are the c-region defining the cleavage site for signal peptidase, residues −6 to −17 are the hydrophobic h-region, and residues −18 to either Met-(−28), Met-(−47), or Met-(−69) are the n-region containing a net positive charge.

The cDNAs for *Drosophila* and *T. marmorata* acetylcholinesterase also have 5′ sequences with multiple ATG sites (2, 3). In contrast to human cholinesterase cDNA, most of these correspond to very brief coding regions that are terminated by stop codons.

In the 3′ untranslated region, a polyadenylylation signal, AATAAA, at nucleotides 2002–2007, is 18 nucleotides away from a poly(A)-addition site, CA, at 2025. This agrees with the 10- to 30-nucleotide separation commonly found for these sites (28). Clone Z3 ends at nucleotide 2027, suggesting that the mRNA for clone Z3 had a poly(A) tail following nucleotide 2026 and that the poly(A) signal at nucleotide 2002 was functional for clone Z3. Cleavage of pre-mRNA prior to polyadenylylation can require the presence of G+T-rich sequences about 30 bp downstream from the CA poly(A) site (28, 29). Fig. 2 shows that GTG, TTTT, and TGT follow nucleotide 2025 within ≈30 bases. This gives additional support to the possibility that transcription termination for the mRNA of clone Z3 was directed by the AATAAA signal at nucleotides 2002–2007.

Clone Z35 terminates at the oligo(A) stretch shown in Fig. 2, at nucleotide 2214. Clone Z35 may have utilized a nonconsensus poly(A)-addition signal, probably ATTAAA (28), at nucleotides 2182–2187. This would suggest termination at alternative poly(A) sites and generation of mRNAs of different lengths but identical coding potential (30).

The amino acid sequence of human cholinesterase is 54% identical with the amino acid sequence of acetylcholinesterase from the electric organ of *T. californica* (5). The homology extends throughout the two proteins, particularly in the region of amino acids 430–450. The sequences of human cholinesterase and *D. melanogaster* acetylcholinesterase (2) are 38% identical. Despite the presence of an active-site serine in the cholinesterases, the cholinesterases do not have significant sequence homology with the serine proteases. It appears that the cholinesterases are a specialized type of serine esterase, differing from serine proteases, which include the blood coagulation factors and other hydrolytic enzymes. To date, the only protein outside of the cholinesterase family that is known to have significant sequence homology with the cholinesterases is thyroglobulin (1, 5).

Note Added in Proof. While this manuscript was being processed, a cDNA sequence based upon two overlapping clones from human fetal brain and fetal liver, corresponding to cholinesterase, was reported by Prody *et al.* (31). Our results are generally in very close agreement, with only a few differences. Our cDNA is longer at the 5′ end by 53 bp. In addition, we have two nucleotide differences at Tyr-(−39) (TAC) and Arg-470 (CGT), which are reported as TGC and CGG, respectively, in ref. 31. Our sequence analysis has 7 fewer nucleotides between bases 2100 and 2185.

The outstanding technical assistance of Patricia T. Brandt is acknowledged. Research was supported by U.S. Army Medical Research and Development Command Contract DAMD17-86-C-6037 (to O.L.), by National Institutes of Health Grant GM27028 (to B.N.L.), and by National Science Foundation Grant DMB-8309705 and grants from the Muscular Dystrophy Association (to T.L.R.). Postdoctoral fellowship support for C.M. was from the Muscular Dystrophy Association and from National Institutes of Health Training Grant NS07118.

1. Schumacher, M., Camp, S., Maulet, Y., Newton, M., MacPhee-Quigley, K., Taylor, S. S., Friedmann, T. & Taylor, P. (1986) *Nature (London)* **319**, 407–409.
2. Hall, L. M. C. & Spierer, P. (1986) *EMBO J.* **5**, 2949–2954.
3. Sikorav, J. L., Krejci, E. & Massoulie, J. (1987) *EMBO J.* **6**, 1865–1873.
4. Prody, C., Zevin-Sonkin, D., Gnatt, A., Koch, R., Zisling, R., Goldberg, O. & Soreq, H. (1986) *J. Neurosci. Res.* **16**, 25–35.
5. Lockridge, O., Bartels, C. F., Vaughan, T. A., Wong, C. K., Norton, S. E. & Johnson, L. L. (1987) *J. Biol. Chem.* **262**, 549–557.
6. Massoulie, J. & Bon, S. (1982) *Annu. Rev. Neurosci.* **5**, 57–106.
7. Vigny, M., Gisiger, V. & Massoulie, J. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2588–2592.
8. Brimijoin, S., Mintz, K. P. & Alley, M. C. (1983) *Mol. Pharmacol.* **24**, 513–520.
9. Hodgkin, W. E., Giblett, E. R., Levine, H., Bauer, W. & Motulsky, A. G. (1965) *J. Clin. Invest.* **44**, 486–493.
10. MacPhee-Quigley, K., Vedvick, T. S., Taylor, P. & Taylor, S. S. (1986) *J. Biol. Chem.* **261**, 13565–13570.
11. Lockridge, O., Adkins, S. & La Du, B. N. (1987) *J. Biol. Chem.* **262**, in press.
12. Kamholz, J., DeFerra, F., Puckett, C. & Lazzarini, R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4962–4966.
13. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
14. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), p. 322.
15. Johnson, D. A., Gautsch, J. W., Sportsman, J. R. & Elder, J. H. (1984) *Gene Anal. Tech.* **1**, 3–8.
16. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
17. Sanger, F., Coulson, A. R., Barrell, B. J., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **13**, 161–178.
18. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55**, 1119–1150.
19. Berger, J., Garattini, E., Hua, J.-C. & Udenfriend, S. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 695–698.
20. Brown, S. S., Kalow, W., Pilz, W., Whittaker, M. & Woronick, C. L. (1981) *Adv. Clin. Chem.* **22**, 1–123.
21. Zakut, H., Matzkel, A., Schejter, E., Avni, A. & Soreq, H. (1985) *J. Neurochem.* **45**, 382–389.
22. Atack, J. R., Perry, E. K., Bonham, J. R., Candy, J. M. & Perry, R. H. (1986) *J. Neurochem.* **47**, 263–277.
23. Haas, R. & Rosenberry, T. L. (1985) *Anal. Biochem.* **148**, 154–162.
24. Haas, R., Brandt, P. T., Knight, J. & Rosenberry, T. L. (1986) *Biochemistry* **25**, 3098–3105.
25. Kozak, M. (1983) *Microbiol. Rev.* **47**, 1–45.
26. Von Heijne, G. (1985) *J. Mol. Biol.* **184**, 99–105.
27. Leff, S. E., Rosenfeld, M. G. & Evans, R. M. (1986) *Annu. Rev. Biochem.* **55**, 1091–1117.
28. Birnstiel, M. L., Busslinger, M. & Strub, K. (1985) *Cell* **41**, 349–359.
29. McLauchlan, J., Gaffney, D., Whitton, J. L. & Clements, J. B. (1985) *Nucleic Acids Res.* **13**, 1347–1368.
30. Parnes, J. R., Robinson, R. R. & Seidman, J. G. (1983) *Nature (London)* **302**, 449–452.
31. Prody, C. A., Zevin-Sonkin, D., Gnatt, A., Goldberg, O. & Soreq, H. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3555–3559.